Published in final edited form as:

J R Stat Soc Series B Stat Methodol. 2015 March 1; 77(2): 397–415. doi:10.1111/rssb.12072.

Semiparametric transformation models for causal inference in time to event studies with all-or-nothing compliance

Wen Yu,

Fudan University, Shanghai, China

Kani Chen,

Hong Kong University of Science and Technology, Hong Kong

Michael E. Sobel, and

Columbia University, New York, USA

Zhiliang Ying

Columbia University, New York, USA

Abstract

We consider causal inference in randomized survival studies with right censored outcomes and all-or-nothing compliance, using semiparametric transformation models to estimate the distribution of survival times in treatment and control groups, conditional on covariates and latent compliance type. Estimands depending on these distributions, for example, the complier average causal effect (CACE), the complier effect on survival beyond time t , and the complier quantile effect are then considered. Maximum likelihood is used to estimate the parameters of the transformation models, using a specially designed expectation-maximization (EM) algorithm to overcome the computational difficulties created by the mixture structure of the problem and the infinite dimensional parameter in the transformation models. The estimators are shown to be consistent, asymptotically normal, and semiparametrically efficient. Inferential procedures for the causal parameters are developed. A simulation study is conducted to evaluate the finite sample performance of the estimated causal parameters. We also apply our methodology to a randomized study conducted by the Health Insurance Plan of Greater New York to assess the reduction in breast cancer mortality due to screening.

Keywords

All-or-nothing compliance; Complier average causal effect; Instrumental variable; Randomized trials; Survival analysis; Semiparametric transformation models

1. Introduction

Randomized experiments are the gold standard for assessing the effect of a treatment on an outcome of interest. Researchers are typically interested in several questions. First, what is

the effect of offering the treatment? This, the “intent to treat” estimand (ITT), measures the “pragmatic effectiveness” of the intervention (Sommer and Zeger, 1991). The ITT (and possibly subgroup variation in this parameter) is of key interest to policy makers, whose primary concern is whether or not the treatment works in practice.

The ITT fails to measure the “efficacy” of the treatment because subjects assigned to the treatment group do not necessarily take up treatment and subjects assigned to the control group may in some instances gain access to treatment outside the experimental context. In earlier work, researchers often performed an “as treated” or “per protocol” analysis to estimate efficacy. Neither analysis contends with the selection bias introduced by subjects’ post-randomization treatment choices.

To remedy this deficiency, a number of authors introduced similar frameworks, leading to estimation of what is now called the complier average causal effect (CACE) (Angrist, Imbens and Rubin, 1996; Bloom, 1984; Baker and Lindeman, 1994; Cuzick, Edwards and Segnan, 1997; Imbens and Angrist, 1994). Here we follow the treatment in Angrist et al. (1996). Briefly, the idea is to split the subjects into subgroups: 1) subjects who always take treatment, whether or not it is assigned (always takers), 2) subjects who take (do not take) treatment when it is (is not) assigned (compliers), 3) subjects who do not take treatment, whether it is assigned or not (never takers), and 4) subjects who take treatment when it is not assigned and who do not take treatment when it is assigned (defiers). The ITT is thus a weighted average of the subgroup ITT’s. Assuming a) a subject’s outcome depends only on his/her treatment assignment (no interference) b) there are no defiers (monotonicity), c) outcomes for always takers (never takers) are the same in both the treatment and control group (exclusion restriction for never takers and always takers), implying the never taker and always taker ITT’s are 0, and d) the complier proportion is non-zero, the CACE is then the ITT divided by the complier proportion.

Subsequent researchers extended the framework in a number of directions and there is now a large literature on estimation and identification of treatment effects in the presence of noncompliance. For example, Imbens and Rubin (1997) established the nonparametric identifiability of the complier control and treatment group distributions under the assumptions above. Cheng, Qin and Zhang (2009a) and Cheng, Small, Tan and Ten Have (2009b) combined instrumental variable methods with empirical likelihood to semiparametrically estimate the complier distributions. Cheng (2009) and Baker (2011) considered outcomes following a multinomial distribution. Little and Yau (1998) included covariates, using maximum likelihood to estimate the distribution of a continuous outcome for never takers under the exclusion restriction and for compliers in the treatment and control groups, assuming normality, conditional on covariates. In their analysis, as is often the case in randomized studies, subjects in the control group did not have access to the treatment, so there are no always takers. Analyzing these same data, Sobel and Muthen (2012) further split the compliers into a class where the treatment effect is 0 and a class with a non-zero effect. Hirano, Imbens, Rubin and Zhou (2000), who also incorporated covariates, took up the case of a binary response and relaxed the never taker and always taker exclusion restriction. While all the papers above considered the case of all-or-nothing compliance, Jin and Rubin (2008) considered the case of partial compliance, while Small,

Ten Have, Joffe and Cheng (2006) considered longitudinal binary data with time varying compliance.

Less attention has been devoted to noncompliance with time to event outcomes. For all-or-nothing compliance, Baker (1998) developed a likelihood based approach for discrete time survival data to estimate the effect (defined as the difference between complier hazards in treatment and control groups) of screening among compliers on breast cancer mortality, taking into consideration censoring by death from competing risks. Frangakis and Rubin (1999) estimated survival curves for compliers in the treatment and control groups, assuming independent censoring within compliance classes. They also assumed that subjects in the control group cannot access treatment, that is, there are no always takers. Robins and Tsiatis (1991) and Mark and Robins (1993) considered a semiparametric “structural” accelerated failure time model with noninformative censoring, relating a subject’s unobserved baseline failure time in the absence of treatment to his/her survival time under treatment received. Again, for all-or-nothing compliance, with no access to treatment outside the treatment arm, Loeys and Goetghebeur (2003) defined and estimated a complier proportional hazards effect of treatment, assuming independent censoring. Cuzick, Sasieni, Myles and Tyrer (2007), who also assumed independent censoring, also considered a proportional hazards model. Both the accelerated failure time models and the proportional hazards model are semiparametric. Gong (2008) considered parametric survival models and estimated complier effects under the assumption of independent censoring within compliance classes. More recently, Nie, Cheng and Small (2011) proposed an estimation procedure based on empirical likelihood for the effect of treatment on survival probabilities among compliers.

This paper also takes up all-or-nothing compliance with time to event outcomes and right censoring. Except for Mark and Robins (1993), Gong (2008), and Cuzick et al. (2007), previous work on this subject has not taken covariates into consideration. This is a serious limitation, as subgroup variation in treatment effects is often of considerable scientific and practical interest. Further, approaches which cannot handle covariates cannot be used to make causal inferences from observational studies, where it is necessary to adjust for confounding. In Mark and Robins (1993), an accelerated failure time model is considered. Gong (2008) considers parametric survival models. Cuzick et al. (2007) assume the event times follow a proportional hazards model. To estimate the model when there are no covariates, they use a generalization of the Mantel-Haenszel estimator based upon ad hoc estimating equations, and for the special case where the covariates are independent of compliance type and their associations with the response are the same across compliance classes, a partial likelihood approach. When covariates and compliance type are not independent but the covariate-outcome relationships are still the same across compliance types, a full likelihood is developed, leading to a complex estimator whose properties were not analytically studied. Unlike a number of the authors above, Cuzick et al. (2007) do not assume the always taker class has zero probability.

Our work is readily seen to extend that of Cuzick et al. (2007). We extend semiparametric linear transformation models, which include the proportional hazards model considered by Cuzick et al. (2007) and the proportional odds model as special cases, and which have

received considerable attention (Cheng, Wei and Ying, 1995; Chen, Jin and Ying, 2002; Zeng and Lin, 2006; 2007), to estimate the distribution of survival times in the treatment and control groups, conditional not only on observed covariates, but on latent compliance type as well. Like Cuzick et al. (2007), we allow for an always taker class with positive probability, but unlike them, we allow the association between covariates and response to vary with compliance class.

We construct a nonparametric maximum likelihood estimator of the model parameters by extending the approach of Zeng and Lin (2007) to accommodate mixture distributions, which result from the fact that compliance types are not fully observed. We analytically establish the asymptotic properties of the estimator, showing that it is consistent and asymptotically normal, with variance reaching the semiparametric efficiency bound. Due to the inclusion of an infinite dimensional parameter in the transformation model and the mixture structure of the likelihood, maximization of the likelihood is computationally challenging, and we use a specially designed expectation-maximization (EM) algorithm to perform the maximization. We then develop inferential procedures for the conditional and unconditional treatment effects of interest.

We proceed as follows. Section 2 introduces the notation, causal framework and causal estimands of interest, the linear transformation models and the assumptions used to identify the causal effects. In Section 3, the observed data likelihood is derived, the EM algorithm used in the maximization is described, and asymptotic properties of the estimators of the model parameters are given. We also derive estimators of the effects from the estimated model and discuss methods for inference. In Section 4, we assess the finite sample performance of these methods in a simulation study and apply our methodology to data from a randomized study (HIP study) conducted by the Health Insurance Plan of Greater New York to assess the cancer mortality reduction due to screening for breast cancer (Shapiro, Venet, Strax and Venet, 1988). Baker (1998), Joffe (2001), Gong (2008) and Nie et al. (2011) also analyzed these data. Section 5 concludes.

2. Notation, assumptions, causal effects and model specification

We consider two arm randomized trials with all-or-nothing compliance and time to event outcomes. For subjects $i = 1, \dots, n$, let $\mathbf{X}_i$ be a p -dimensional covariate vector. Let $\mathbf{Z} = (Z_1, \dots, Z_n)$ denote the n -vector of treatment assignments, taking values of 0 and 1: $Z_i = 0$ if i is assigned to treatment group 0 (hereafter referred to as the control group), $Z_i = 1$ if i is assigned to treatment group 1 (hereafter referred to as the treatment group), with $0 < P(Z_i = 1) < 1$. For all possible assignment vectors $\mathbf{z} = (z_1, \dots, z_n) \in \{0, 1\}^n$, let $D_i^{\mathbf{z}} = 1$ if, under assignment $\mathbf{z}$, i takes up the treatment offered in the treatment group, 0 otherwise. Similarly, let $T_i^{\mathbf{z}}$ and $C_i^{\mathbf{z}}$ denote, respectively, the potential event time and potential right censoring time for subject i under assignment $\mathbf{z}$.

A.1 Stable unit treatment value assumption

(SUTVA, Rubin, 1980). For any assignments $\mathbf{z}$ and $\mathbf{z}'$, if $z_i = z'_i$ then $D_i^{\mathbf{z}} = D_i^{\mathbf{z}'}$, $T_i^{\mathbf{z}} = T_i^{\mathbf{z}'}$ and $C_i^{\mathbf{z}} = C_i^{\mathbf{z}'}$ for $i = 1, \dots, n$. Thus, $D_i^{\mathbf{z}} \equiv D_i^{z_i} \equiv D_i^z$, where $z = 0$ if $z_i = 0$, 1 otherwise; similarly, $T_i^{\mathbf{z}} \equiv T_i^{z_i} \equiv T_i^z$, $C_i^{\mathbf{z}} \equiv C_i^{z_i} \equiv C_i^z$.

A.2 Random sampling

$(D_i^z, T_i^z, C_i^z, \mathbf{X}_i, Z_i)$, $i = 1, \dots, n$, are independent and identically distributed (i.i.d.) copies from the distribution of a random vector $(D^z, T^z, C^z, \mathbf{X}, Z)$.

A.3 Random assignment conditional on covariates

$D^0, T^0, C^0, D^1, T^1, C^1 \perp\!\!\!\perp Z \mid \mathbf{X}$, where the symbol $\perp\!\!\!\perp$ denotes independence.

In a completely randomized experiment, the stronger statement $D^0, T^0, C^0, D^1, T^1, C^1, \mathbf{X} \perp\!\!\!\perp Z$ holds. The weaker assumption A.3 allows consideration of conditionally randomized experiments and observational studies with confounders $\mathbf{X}$.

Let $U_i = k$ denote i 's latent class membership: $U_i = 1$ if i is an always taker ($D_i^0 = D_i^1 = 1$), 2 if i is a complier ($D_i^0 = 0, D_i^1 = 1$), 3 if i is a never taker ($D_i^0 = D_i^1 = 0$), 4 if i is a defier ($D_i^0 = 1, D_i^1 = 0$).

A.4 Conditional non-null compliance class

$$P(U = 2 \mid \mathbf{X} = \mathbf{x}) > 0.$$

A.5 Conditional monotonicity

$$P(D^1 \geq D^0 \mid \mathbf{X} = \mathbf{x}) = 1.$$

A.6 Exclusion restriction

For $k = 1, 3$ and for all t , $P(T^0 = t \mid U = k, \mathbf{X} = \mathbf{x}) = P(T^1 = t \mid U = k, \mathbf{X} = \mathbf{x})$.

Assumption A.4 states that the complier class has non-zero probability at every covariate level, while assumption A.5 implies the defier class has probability zero. Because always takers and never takers receive the same treatment regardless of assignment, the experiment provides no information about the treatment effect in these classes. Assumption A.6 additionally rules out a “direct effect” of treatment assignment on the outcome in these classes.

Three estimands are discussed explicitly in this paper: the (conditional) complier average causal effect CACE($\mathbf{x}$)

$$E(T^1 - T^0 \mid U=2, \mathbf{X}=\mathbf{x}), \quad (1)$$

the time t effect on the (conditional) complier survival probability CESP($t; \mathbf{x}$)

$$P(T^1 > t | U=2, \mathbf{X}=\mathbf{x}) - P(T^0 > t | U=2, \mathbf{X}=\mathbf{x}), \quad (2)$$

and the (conditional) complier quantile causal effect $CQCE(q; \mathbf{x})$

$$\sup\{t: P(T^1 \leq t | U=2, \mathbf{X}=\mathbf{x}) \leq q\} - \sup\{t: P(T^0 \leq t | U=2, \mathbf{X}=\mathbf{x}) \leq q\}. \quad (3)$$

Note that $CACE(\mathbf{x})$ may not exist and may not be identified. Unconditional versions of these estimands are denoted as $CACE$, $CESP(t)$ and $CQCE(q)$. Gong (2008) and Nie et al. (2011) considered $CESP(t)$. Cheng et al. (2009a) considered $CACE$ and $CQCE(q)$.

The estimands above are functions of the potential event time distributions. To model these distributions we propose the following causal linear transformation model:

$$\log H(T_i^z) = - \sum_{k=1}^3 (\beta_{0k}^c + \beta_k^c z + \eta_k^{c\top} \mathbf{X}_i) I(U_i=k) + \varepsilon_i^z, \quad (4)$$

where $H(\cdot)$ is an unspecified continuously differentiable increasing function with $H(0) = 0$; for $k = 1, 2, 3$, β_{0k}^c is a within class intercept, β_k^c is a within class average effect of treatment assignment on the log H scale, η_k^c is a within class parameter vector relating covariates to the potential outcomes, $I(\cdot)$ is the indicator function, and for $z = 0, 1$, the random errors $\varepsilon_i^z \perp U_i, \mathbf{X}_i$ both following a known distribution F_ε . One typical choice for F_ε is the class of logarithmic transformations $F_\varepsilon(x) = 1 - \{1 + r \exp(x)\}^{-1/r}$ for $r > 0$ and $F_\varepsilon(x) = 1 - \exp\{-\exp(x)\}$ for $r = 0$, with $r = 1$ corresponding to the proportional odds model and $r = 0$ to the proportional hazards model. To uniquely determine the parameters, we treat compliers assigned to the control group as the reference group, i.e., setting $\beta_{02}^c = 0$. Also, under the exclusion restriction A.6, $\beta_k^c = 0$ for $k = 1, 3$. The special case where subjects cannot access treatment outside of the control group, i.e., there are no always takers, is easily accommodated by requiring $k = 1$, and if all subjects are compliers, by requiring $k = 1, 3$. Although the parameters η_k^c do not depend on treatment assignment, this restriction is not necessary.

As only (in the absence of censoring) the event time $T_i = Z_i T_i^1 + (1 - Z_i) T_i^0$ is observed, as opposed to both potential outcomes T_i^0 and T_i^1 , the causal model (4) cannot be estimated directly using the observed data. We can only estimate the parameters of the analogous regression model relating observed event times T_i to compliance type, covariates and assignment:

$$\log H(T_i) = - \sum_{k=1}^3 (\beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i) I(U_i=k) + \varepsilon_i, \quad (5)$$

where the random errors ε_i are i.i.d. with distribution F_ε and are independent of $U_i, \mathbf{X}_i, Z_i$. As above, the parameters η_k do not depend on treatment assignment, although the estimation procedure proposed in Section 3 can easily accommodate this generalization. If there are no always takers, $k = 1$. If all subjects are compliers $k = 1, 3$; in this case, the model (5) reduces

to $\log \log H(T_i) = -\beta_2 Z_i - \eta_2^\top \mathbf{X}_i + \varepsilon_i$, the usual linear transformation model for survival data (Cheng et al., 1995; Chen et al., 2002; Zeng and Lin, 2007).

In general, analogous parameters in the regression model (5) and the causal model (4) are not equal; thus the causal estimands (1)–(3) and their unconditional counterparts, which depend on the parameters in the causal model (4), are not in general identified from the regression model (5). However, under the randomization assumption A.3, the parameters of the causal model (4) are equal to the analogous parameters of (5): for any t , k , $\mathbf{x}$ and z ,

$$\begin{aligned}
 & F_\varepsilon\{\log H(t) + \beta_{0k}^c + \beta_k^c z + \eta_k^{c\top} \mathbf{x}\} \\
 &= P\{\varepsilon^z \leq \log H(t) \\
 & \quad + \beta_{0k}^c + \beta_k^c z + \eta_k^{c\top} \mathbf{x} | U \\
 &= k, \mathbf{X}\} \\
 &= \mathbf{x}\} = P(T^z \leq t | U=k, \mathbf{X}=\mathbf{x}) = P(T \leq t | U=k, \mathbf{X}=\mathbf{x}, Z=z) = P\{\varepsilon \leq \log H(t) \\
 & \quad + \beta_{0k}^c + \beta_k^c z + \eta_k^{c\top} \mathbf{x} | U \\
 &= k, \mathbf{X}\} \\
 &= \mathbf{x}, Z \\
 &= z\} = F_\varepsilon\{\log H(t) \\
 & \quad + \beta_{0k}^c + \beta_k^c z + \eta_k^{c\top} \mathbf{x}\},
 \end{aligned} \tag{6}$$

resulting in $\beta_{0k}^c = \beta_{0k}$, $\beta_k^c = \beta_k$ and $\eta_k^c = \eta_k$ for $k = 1, 2, 3$. Therefore, when (6) holds, as in a randomized experiment, the regression model (5) can be used to estimate (1)–(3) and their unconditional counterparts under additional conditions discussed in Section 3.

For the censoring scheme, we assume the potential outcomes are independently censored:

A.7 Independent censoring of potential outcomes

For $z = 0, 1$, $C^z \perp\!\!\!\perp \varepsilon^z$; $C^z \perp\!\!\!\perp U | \mathbf{X}$.

Assumption A.7 implies $C^z \perp\!\!\!\perp T^z, U | \mathbf{X}$. Note that our assumption does not exclude the case where the potential censoring time depends on treatment assignment. For subject i , the observed censoring time $C_i = Z_i C_i^1 + (1 - Z_i) C_i^0$. If treatment assignment does not affect the censoring time, $C_i^0 = C_i^1 = C_i$. As an example, under Type I censoring at some fixed time c , $C_i^0 = C_i^1 = C_i = c$ for all i . Assumptions A.3 and A.7 imply $C \perp\!\!\!\perp T, U | (\mathbf{X}, Z)$. When the data are right censored, we observe the random variables $Y_i = \min\{T_i, C_i\}$ and $\delta_i = I\{T_i < C_i\}$, $i = 1, \dots, n$. Let $D_i = Z_i D_i^1 + (1 - Z_i) D_i^0$ be the treatment received. The observed data are thus $(Y_i, \delta_i, \mathbf{X}_i, D_i, Z_i)$, $i = 1, \dots, n$, with realized values $(y_i, \delta_i, \mathbf{x}_i, d_i, z_i)$, $i = 1, \dots, n$.

Because compliance status is only partially observed, the log-likelihood of the observed data is a mixture of distributions depending on the compliance type probabilities $p_{ik} = P(U_i = k | \mathbf{X}_i = \mathbf{x}_i)$, $i = 1, \dots, n$, $k = 1, 2, 3$. In the following, we assume that the p_{ik} follow a parametric regression model with a finite dimensional parameter vector θ . The particular form we use is the multinomial logit model $\log\{p_{ik}(\theta)/p_{i2}(\theta)\} = \theta_k^\top \mathbf{x}_i$, $k = 1, 3$ and $\theta = (\theta_1^\top, \theta_3^\top)^\top$.

3. Estimation

In this section, maximum likelihood is used to estimate the unknown parameters of model (5). In Section 3.1, the likelihood is derived, and in Section 3.2, the EM algorithm used to maximize the likelihood is presented. Section 3.3 discusses asymptotic properties of the estimators (with proofs in Appendix A.2). Section 3.4 discusses inferential procedures for the casual estimands defined in Section 2.

3.1. Likelihood construction

Let $f_\varepsilon(t) = F'_\varepsilon(t)$ be the error density, $\lambda_\varepsilon(t) = f_\varepsilon(t)/\{1 - F_\varepsilon(t)\}$ the hazard function and $\Lambda_\varepsilon(t) = \int_{-\infty}^t \lambda_\varepsilon(s) ds$ the cumulative hazard function. Let $\vartheta = (\vartheta^{par}, H)$ be the collection of all unknown parameters, where $\vartheta^{par} = (\theta^\top, \beta_{01}, \beta_2, \beta_{03}, \eta_1^\top, \eta_2^\top, \eta_3^\top)^\top$ is the parametric part, and $\mathcal{O}_i = (y_i, \delta_i, \mathbf{x}_i, d_i, z_i)$ the observed data for subject i . For $i = 1, \dots, n$ and $k = 1, 2, 3$, let

$$g_{ik} = \left[\lambda_\varepsilon \{ \log H(y_i) + \beta_{0k} + \beta_k z_i + \eta_k^\top \mathbf{x}_i \} \frac{h(y_i)}{H(y_i)} \right]^{\delta_i} \times \exp \{ -\Lambda_\varepsilon \{ \log H(y_i) + \beta_{0k} + \beta_k z_i + \eta_k^\top \mathbf{x}_i \} \},$$

where $h(t) = H'(t)$. Under (5) and the assumptions in Section 2, the joint density of $(Y_i, \delta_i, D_i, U_i, \mathbf{X}_i, Z_i)$ is

$$\begin{aligned} & f_{Y, \Delta, D, U, \mathbf{X}, Z}(y_i, \delta_i, d_i, u_i, \mathbf{x}_i, z_i; \vartheta) \\ &= f_{\mathbf{X}, Z}(\mathbf{x}_i, z_i) \\ & \times \left\{ \sum_{k=1}^3 p_{ik} g_{ik} \lambda_{C|D, U, \mathbf{X}, Z}^{1-\delta_i}(y_i | d_i, u_i, \mathbf{x}_i, z_i) S_{C|D, U, \mathbf{X}, Z}(y_i | d_i, u_i, \mathbf{x}_i, z_i) I(u_i = k) \right\}, \end{aligned}$$

where $f_{\mathbf{X}, Z}$ is the density of $(\mathbf{X}, Z)$ and $\lambda_{C|D, U, \mathbf{X}, Z}$ and $S_{C|D, U, \mathbf{X}, Z}$ are the conditional hazard and survivor functions of C given $(D, U, \mathbf{X}, Z)$. As $C \perp\!\!\!\perp U | (\mathbf{X}, Z)$, $\lambda_{C|D, U, \mathbf{X}, Z}(y_i | d_i, u_i, \mathbf{x}_i, z_i) = \lambda_{C|\mathbf{X}, Z}(y_i | \mathbf{x}_i, z_i)$ and $S_{C|D, U, \mathbf{X}, Z}(y_i | d_i, u_i, \mathbf{x}_i, z_i) = S_{C|\mathbf{X}, Z}(y_i | \mathbf{x}_i, z_i)$. Integrating out the unobserved variable U_i gives the likelihood function for the observed data $(y_i, \delta_i, d_i, \mathbf{x}_i, z_i)$, $i = 1, \dots, n$:

$$\begin{aligned} & \prod_{i=1}^n \{ (p_{i1}g_{i1} + p_{i2}g_{i2})I(d_i = z_i = 1) + p_{i3}g_{i3}I(d_i = 0, z_i = 1) + p_{i1}g_{i1}I(d_i = 1, z_i = 0) + (p_{i2}g_{i2} + p_{i3}g_{i3})I(d_i = z_i = 0) \} \\ & \times \lambda_{C|\mathbf{X}, Z}^{1-\delta_i}(y_i | \mathbf{x}_i, z_i) S_{C|\mathbf{X}, Z}(y_i | \mathbf{x}_i, z_i) f_{\mathbf{X}, Z}(\mathbf{x}_i, z_i). \end{aligned}$$

Under the usual assumption that the distributions of the censoring time and covariates do not depend on the parameters of interest, the kernel of the likelihood (omitting components not involving the parameters of interest) is given by

$$L^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta) = \prod_{d_i = z_i = 1} (p_{i1}g_{i1} + p_{i2}g_{i2}) \prod_{d_i = 0, z_i = 1} p_{i3}g_{i3} \times \prod_{d_i = 1, z_i = 0} p_{i1}g_{i1} \prod_{d_i = z_i = 0} (p_{i2}g_{i2} + p_{i3}g_{i3}).$$

Following Zeng and Lin (2006; 2007), we maximize L^O with respect to ϑ directly; the solution for $H(\cdot)$ is a right continuous step function with jumps only on the uncensored observed times. Specifically, let $y_{(1)}, \dots, y_{(n_u)}$ denote the ordered uncensored observations, where n_u is the number of uncensored observations. We restrict $H(\cdot)$ to be an increasing step function on $[0, \infty)$, starting from 0 and having jumps at $y_{(1)}, \dots, y_{(n_u)}$. Let h_j be the jump size at $y_{(j)}$ for $j = 1, \dots, n_u$. Replace $h(y_{(j)})$ and $H(y_i)$ in L^O by h_j and $\sum_{j=1}^{n_u} y_i h_j$, respectively. Then the infinite dimensional parameter $H(\cdot)$ becomes an n_u -dimensional parameter $(h_1, \dots, h_{n_u})$. To obtain the maximum likelihood estimator (MLE), we maximize L^O with respect to $(\theta, \beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3, h_1, \dots, h_{n_u})$, subject to the constraints that $h_1, \dots, h_{n_u}$ are non-negative. Denote the maximizer by $(\hat{\theta}, \hat{\beta}_{01}, \hat{\beta}_2, \hat{\beta}_{03}, \hat{\eta}_1, \hat{\eta}_2, \hat{\eta}_3, \hat{h}_1, \dots, \hat{h}_{n_u})$. Write

$\hat{\vartheta}^{par} = (\hat{\theta}^\top, \hat{\beta}_{01}, \hat{\beta}_2, \hat{\beta}_{03}, \hat{\eta}_1^\top, \hat{\eta}_2^\top, \hat{\eta}_3^\top)^\top$ and $\hat{H}(t) = \sum_{j=1}^{n_u} t \hat{h}_j$ for $t \geq y_{(j)}$ while $\hat{H}(t) = 0$ for $t < y_{(1)}$. The MLE of ϑ is defined to be $\hat{\vartheta} = (\hat{\vartheta}^{par}, \hat{H})$.

3.2. Algorithm

Maximization of the likelihood L^O is difficult because the dimension of ϑ (throughout this subsection $\vartheta = (\theta, \beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3, h_1, \dots, h_{n_u})$) increases as the sample size increases and the components of the likelihood are mixtures of distributions over the latent classes. To obtain $\hat{\vartheta}$, we use the following EM algorithm.

E-step: Let $\vartheta^{(0)}$ be the starting value for ϑ . We calculate the conditional expectation at $\vartheta = \vartheta^{(0)}$ of the log-likelihood given observations

$$\sum_{i=1}^n E_{\vartheta^{(0)}} \left[\log \{ f_{Y, \Delta, D, U, X, Z}(Y_i, \Delta_i, D_i, U_i, \mathbf{X}_i, Z_i; \vartheta) \} \middle| \mathcal{O}_i \right] \propto \sum_{i=1}^n \sum_{k=1}^3 \{ \log(p_{ik}) + \log(g_{ik}) \} r_{ikd_i z_i}(\vartheta^{(0)}),$$

where $r_{ikd_i z_i}(\vartheta) = P_{\vartheta}(U_i = k | Y_i = y_i, \Delta_i = \delta_i, D_i = d_i, \mathbf{X}_i = \mathbf{x}_i, Z_i = z_i)$, $i = 1, \dots, n$, $k = 1, 2, 3$.

A careful calculation gives: $r_{ik11}(\vartheta) = p_{ik} g_{ik} / (p_{i1} g_{i1} + p_{i2} g_{i2})$ for $k = 1, 2$ and 0 for $k = 3$; $r_{ik01}(\vartheta) = 0$ for $k = 1, 2$ and 1 for $k = 3$; $r_{ik10}(\vartheta) = 1$ for $k = 1$ and 0 for $k = 2, 3$, and $r_{ik00}(\vartheta) = p_{ik} g_{ik} / (p_{i2} g_{i2} + p_{i3} g_{i3})$ for $k = 2, 3$ and 0 for $k = 1$. Finally, denote

$$\sum_{i=1}^n \sum_{k=1}^3 \{ \log(p_{ik}) + \log(g_{ik}) \} r_{ikd_i z_i}(\vartheta^{(0)}) \text{ by } l_E^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}).$$

M-step: Update the parameter value by taking

$$\vartheta^{(1)} = \arg \max_{\vartheta} l_E^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}),$$

subject to the constraints that $h_1, \dots, h_{n_u}$ are non-negative.

The E-step is repeated using $\vartheta^{(1)}$, and iteration continues until convergence.

It is easy to see that

$$l_E^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}) = l_{E,1}^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}) + l_{E,2}^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}),$$

where $l_{E,1}^O(\theta_1, \dots, \theta_n; \vartheta, \vartheta^{(0)}) = \sum_{i=1}^n \sum_{k=1}^3 \log(p_{ik}) r_{ikd_i z_i}(\vartheta^{(0)})$ and $l_{E,2}^O(\theta_1, \dots, \theta_n; \vartheta, \vartheta^{(0)}) = \sum_{i=1}^n \sum_{k=1}^3 \log(g_{ik}) r_{ikd_i z_i}(\vartheta^{(0)})$. The maximization of $l_{E,1}^O$ involves only θ , while the maximization of $l_{E,2}^O$ involves the remaining parameters. This separation is important. First, maximizing $l_{E,1}^O$ is usually easy for the parametric models commonly chosen for the p_{ik} 's. Second, note that at each iteration, $r_{ikd_i z_i}(\vartheta_0)$ is treated as a constant, so $l_{E,2}^O$ is a weighted sum of $\log g_{ik}$'s. Consequently, the maximization of $l_{E,2}^O$ is similar to the maximization of the likelihood for right censored data under the usual linear transformation models, as discussed in Zeng and Lin (2006; 2007). We show in Appendix A.1 that when the random error follows the extreme value distribution, that is, the model (5) is the proportional hazards model, the parameters $h_1, \dots, h_{n_H}$ can be profiled out and the maximization of $l_{E,2}^O$ reduces to the maximization of a weighted log-partial likelihood. The weighted log-partial likelihood is concave in $(\beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3)$.

3.3. Large sample properties

Under suitable regularity conditions, the MLE $\hat{\vartheta}$ has the usual large sample properties, including consistency and asymptotic normality, as given in theorems 3.1 and 3.2 below. Let ϑ_0^{par} and $H_0(\cdot)$ denote the true values of ϑ^{par} and $H(\cdot)$, respectively, $\|\cdot\|$ the Euclidean norm, and τ the duration of the study.

Theorem 3.1. *Under conditions C.1–C.5 in Appendix A.2,*

$$\|\hat{\vartheta}^{par} - \vartheta_0^{par}\| + \sup_{t \in [0, \tau]} |\hat{H}(t) - H_0(t)|$$

converges to 0 almost surely.

Theorem 3.2. *Under conditions C.1–C.5 in Appendix A.2, $\sqrt{n}\{\hat{\vartheta}^{par} - \vartheta_0^{par}, \hat{H}(\cdot) - H_0(\cdot)\}$ converges weakly to a zero-mean Gaussian process in the Banach space $\mathbb{R}^m \times l^\infty(\mathcal{D})$, where m is the dimension of $\mathcal{P}^{par}$ and $\mathcal{D}$ is the unit ball in the space of functions on $[0, \tau]$ with bounded variation.*

The proofs of the theorems are given in Appendix A.2. By using arguments similar to those in Zeng and Lin (2006), one can show that $\hat{\vartheta}^{par}$ is semiparametrically efficient. The limiting variances and covariances of the estimators can be estimated using the inverse of the observed information matrix derived from L^O evaluated at $\hat{\vartheta}$. Moreover, we can estimate the limiting variance for any smooth functional of $\hat{\vartheta}$ by the delta method.

3.4. Model-based inference for the complier effects

As previously noted, the causal estimands (1)–(3) and their unconditional counterparts depend on the potential outcome distributions. Under a parametric model, where the entire conditional outcome distribution is specified, $\text{CESP}(t; \mathbf{x})$ is identified for all t , and $\text{CQCE}(q; \mathbf{x})$ is identified for all q under the assumptions previously made; $\text{CACE}(\mathbf{x})$ is also identified

under these assumptions when it exists. In the linear transformation models, only that portion of the conditional outcome distribution within the study duration is identifiable. Thus, if the upper limit of the support of T (at $\mathbf{x}$) is beyond τ , $\text{CACE}(\mathbf{x})$ is not identifiable; similarly, if $A = \{\mathbf{x}: \text{CACE}(\mathbf{x}) \text{ is not identifiable}\}$ and $P(\mathbf{X} \in A \mid U = 2) > 0$, CACE is not identifiable. $\text{CESP}(t; \mathbf{x})$ and $\text{CESP}(t)$ are not identifiable for $t > \tau$, and $\text{CQCE}(q; \mathbf{x})$ and $\text{CQCE}(q)$ may not be identifiable for all q . Here we describe inferential procedures for the estimands when they are identifiable.

Let $F_1(t; \mathbf{x}) = P(T^1 \leq t \mid U = 2, \mathbf{X} = \mathbf{x})$ and $F_0(t; \mathbf{x}) = P(T^0 \leq t \mid U = 2, \mathbf{X} = \mathbf{x})$ for $t \geq 0$. Hence $\text{CACE}(\mathbf{x}) = \int_0^\infty t d\{F_1(t; \mathbf{x}) - F_0(t; \mathbf{x})\}$, $\text{CESP}(t; \mathbf{x}) = \{1 - F_1(t; \mathbf{x})\} - \{1 - F_0(t; \mathbf{x})\} = F_0(t; \mathbf{x}) - F_1(t; \mathbf{x})$, and $\text{CQCE}(q; \mathbf{x}) = F_1^{-1}(q; \mathbf{x}) - F_0^{-1}(q; \mathbf{x})$, where $F_z^{-1}(q; \mathbf{x}) = \sup\{t: F_z(t; \mathbf{x}) \leq q\}$, $z=0, 1$.

Under the assumptions in Section 2, $F_z(t; \mathbf{x}) = F_z\{\log H(t) + z\beta_2 + \eta_2^\top \mathbf{x}\}$. Let $\hat{F}_z(t; \mathbf{x})$ be the plug-in estimator, with H , β_2 and η_2 respectively replaced by $\hat{H}$, $\hat{\beta}_2$ and $\hat{\eta}_2$. Using the MLE of the unknown parameters in the model (5), $\text{CACE}(\mathbf{x})$ can be estimated by the plug-in estimator

$$\hat{\text{CACE}}(\mathbf{x}) = \sum_{i=1}^{n_u} y_{(i)} \{\Delta \hat{F}_1(y_{(i)}; \mathbf{x}) - \Delta \hat{F}_0(y_{(i)}; \mathbf{x})\},$$

where, for $z = 0, 1$, $\hat{F}_z(\hat{y}_{(1)}; \mathbf{x}) = F_z(\hat{y}_{(1)}; \mathbf{x})$ and, for $i \geq 2$, $\hat{F}_z(\hat{y}_{(i)}; \mathbf{x}) = F_z(\hat{y}_{(i)}; \mathbf{x}) - F_z(\hat{y}_{(i-1)}; \mathbf{x})$. Similarly, $\hat{\text{CESP}}(t; \mathbf{x}) = \hat{F}_0(t; \mathbf{x}) - \hat{F}_1(t; \mathbf{x})$ and $\hat{\text{CQCE}}(q; \mathbf{x}) = \hat{F}_1^{-1}(q; \mathbf{x}) - \hat{F}_0^{-1}(q; \mathbf{x})$.

By transformation invariance, $\hat{\text{CACE}}(\mathbf{x})$, $\hat{\text{CESP}}(t; \mathbf{x})$ and $\hat{\text{CQCE}}(q; \mathbf{x})$ are maximum likelihood estimators. Inferences for $\text{CACE}(\mathbf{x})$ and $\text{CESP}(t; \mathbf{x})$ can be obtained by applying the delta method without much difficulty since they are smooth functionals of $\mathfrak{F}$. For $\text{CQCE}(q; \mathbf{x})$, as $\hat{F}_z(t; \mathbf{x})$, $z = 0, 1$, are discontinuous in t , the delta method is no longer applicable. We may use the bootstrap method to obtain the $(1 - \alpha) \times 100\%$ confidence interval for $\text{CQCE}(q; \mathbf{x})$.

The unconditional estimands CACE and $\text{CESP}(t)$ are obtained by averaging $\text{CACE}(\mathbf{x})$ and $\text{CESP}(t; \mathbf{x})$, respectively, with respect to the conditional distribution of $\mathbf{X}$ given $U = 2$. The corresponding estimators can be obtained using the plug-in principle. To estimate the conditional distribution, Bayes rule is used. For $i = 1, \dots, n$, define $w_i = p_{i2}(\hat{\theta})/\bar{p}_2$, where

$$\bar{p}_2 = n^{-1} \sum_{i=1}^n p_{i2}(\hat{\theta}).$$

We use $\hat{\text{CACE}} = n^{-1} \sum_{i=1}^n w_i \hat{\text{CACE}}(\mathbf{x}_i)$ and

$\hat{\text{CESP}}(t) = n^{-1} \sum_{i=1}^n w_i \hat{\text{CESP}}(t; \mathbf{x}_i)$ to estimate CACE and $\text{CESP}(t)$, respectively. To estimate $\text{CQCE}(q)$, we first estimate $P(T^z \leq t \mid U = 2)$ by $\hat{F}_z(t)$, $z = 0, 1$, where

$$\hat{F}_z(t) = n^{-1} \sum_{i=1}^n w_i \hat{F}_z(t; \mathbf{x}_i).$$

The estimator is then defined as:

$$\hat{\text{CQCE}}(q) = \hat{F}_1^{-1}(q) - \hat{F}_0^{-1}(q), \text{ where } \hat{F}_z^{-1}(q) = \sup\{t: \hat{F}_z(t) \leq q\}.$$

It follows from the asymptotic properties of the MLE that the three estimators are consistent and asymptotically normal. Asymptotic confidence intervals for CACE and CESP(t) can be obtained by the delta method, and again, by the bootstrap for CQCE(q).

4. Numerical studies

4.1. Simulation study

We conduct a simulation study to assess the finite sample performance of the proposed inferential procedures. For the model (5), we set $H(t) = t$, $\beta_{01} = -0.5$, $\beta_2 = -2$ and $\beta_{03} = 1$. A one-dimensional covariate X_i uniformly distributed on $[0, 2]$ is generated. We set $\eta_1 = 0.3$, $\eta_2 = 0$, and $\eta_3 = -0.3$. The distribution of ε_i is chosen from the class of logarithmic transformations with $r = 0, 0.5$ and 1 . Note that $r = 0$ and $r = 1$ correspond to the proportional hazards model and the proportional odds model, respectively. The compliance class probabilities follow the multinomial logit model $\log(p_{ik}/p_{i2}) = \theta_k X_i$ for $k = 1, 3$, with $\theta_1 = 1$ and $\theta_3 = -1.2$. Under the model, at the average value of the covariate, 22% of the subjects are always takers, 60% are compliers and 18% are never takers. Z_i , $i = 1, \dots, n$ are i.i.d. Bernoulli random variables with $P(Z_i = 1) = 0.5$. The censoring time C_i is uniform on $[0, \tau]$, where $\tau = 13$ for $r = 0$, $\tau = 20$ for $r = 0.5$ and $\tau = 30$ for $r = 1$, resulting in 20% censoring under each of the models. The sample sizes n are 100 and 200. Under each scenario, 1,000 replications are carried out. For the optimization in the proposed EM algorithm, several functions designed for non-linear optimization in the ‘Optimization Toolbox’ of MATLAB are used. To maximize $l_{E,1}^O$, we use the ‘*fminsearch*’ function, which uses a simplex search method. To maximize $l_{E,2}^O$ subject to $h_1, \dots, h_{n_u}$ being non-negative, we use the ‘*fmincon*’ function, which finds optimal values for nonlinear functions subject to upper and lower bounds constraints. The ‘*fmincon*’ function with the large scale option uses the trust-region-reflective algorithm, while the function with the medium scale option uses the active set algorithm. To obtain the inverse of the observed information matrix, we use a MATLAB function ‘*hessian*’ (associated with some other functions) to calculate the negative Hessian matrix of the observed log-likelihood function $\log L^O$ evaluated at $\hat{\vartheta}$, from which the estimated standard errors are then obtained.

Table 1 assesses the behavior of the parameter estimates. For all the parameters, the average of the estimates is reasonably close to the parameter value, and in general, closer for $n = 200$, as to be expected. The column labeled SE reports the empirical standard error of the estimates. The column labeled SEE is the average of the standard errors estimated using the inverse of the observed information matrix. In almost every instance, SEE is slightly smaller than SE, with SEE closer to SE for $n = 200$ than for 100. Finally, the column labeled CP gives the empirical coverage probability for the Wald-type confidence intervals; these are quite close to the nominal levels, especially for $n = 200$.

We compute $\widehat{\text{CESP}}(t; x)$ with $x = 1$, the average value of the covariate, and $\widehat{\text{CESP}}(t)$ for $t = 1, 2, 3, 4$, using the delta method to obtain the standard errors of $\widehat{\text{CESP}}(t; 1)$ and $\widehat{\text{CESP}}(t)$ and construct the Wald-type confidence intervals. The results are summarized in Table 2.

The averages of the estimates are reasonably close to true values, for all r , for both $n = 100$ and $n = 200$. For all t considered, estimated values of $\text{CESP}(t; 1)$ and $\text{CESP}(t)$ are close, reflecting that $\text{CESP}(t; 1) = \text{CESP}(t)$, as $\eta_2 = 0$. The averages of the estimated standard errors obtained by the delta method (SEE) are in general close to the corresponding standard errors (SE). In all cases, the empirical coverage rate of the Wald-type intervals is slightly smaller than the nominal value. For $n = 200$, the coverage rates are closer to the nominal values.

In survival analysis, the difference in median survival times under treatment versus no treatment is an important parameter of interest. This estimand does not take non-compliance into account; the analogous parameter that adjusts for non-compliance is the complier median effect $\text{CQCE}(0.5)$. In Table 3, the simulated averages and standard errors (SE) of the conditional complier median effect $\hat{\text{CQCE}}(0.5; 1)$ and its unconditional counterpart $\hat{\text{CQCE}}(0.5)$ are reported for the proportional hazards model with $n = 100$ and 200 . The estimated standard errors (SEE) are obtained by the bootstrap method with 100 bootstrap samples for each replication. For both $\text{CQCE}(0.5; 1) = \text{CQCE}(0.5)$, the averages of the estimates are close to the true values, for both $n = 100$ and 200 . Also, the estimated standard errors are close to the corresponding standard errors and the coverage probabilities (CP) based on the Wald-type confidence intervals are reasonable.

4.2. The HIP study

The Health Insurance Plan of Greater New York initiated a randomized study in 1963 to assess the reduction due to screening in breast cancer mortality among women 40–64. A total of 60,696 eligible women were included in this study. Among them, 30,131 women were assigned to the treatment group, which received four yearly screenings, each consisting of a mammogram and physical exam, while the remaining 30,565 women were assigned to usual care. Among the women assigned to the treatment group, almost one-third (9,984) refused treatment after randomization. None of the controls received screening, i.e., there are no always takers. In keeping with previous analyses of these data, compliance is treated as all-or-nothing.

In a previous analysis of data from the HIP study, Baker (1998) considered only subjects whose breast cancer was diagnosed within 7 years of follow up. Adjusting for non-compliance, he estimated the cost effectiveness of screening to be \$16,000 per life year saved. Joffe (2001) also analyzed data from this study, assuming an end of follow up at 10 years. Using a Weibull model, he estimated a 15% increase in the mean time to death from breast cancer in the screening group, and after adjusting for non-compliance, a 22% increase among compliers under screening. Gong (2008), using a 7 year follow up for all subjects and a Weibull model for both censoring and failure times with non-ignorable censoring, estimated $\hat{\text{CESP}}(5) = 0.00148$, with standard error 0.00038; she obtained similar results when non-informative censoring was assumed ($\hat{\text{CESP}}(5) = 0.0018$, with standard error 0.00035). Nie et al. (2011), following both Baker (1998) and Joffe (2001), only included subjects whose breast cancer was diagnosed within 7 years of follow up, and they assumed an end of follow up at 10 years. They estimated $\text{CESP}(1) - \text{CESP}(10)$, reporting

$\hat{CESP}(10)=0.123$ with 95% confidence interval (0.041, 0.206); they stated that screening shows benefits from 4 years on.

We also consider the estimand $CESP(t)$. We use the same sample selection criteria as Nie et al. (2011). The endpoint is death from breast cancer. In these data, the age at randomization for each subject was reported. We use this in the analysis as the covariate X and we additionally consider $CESP(t; x)$, with x standing for the value of X ; note that none of the authors above considered covariates when analyzing the data.

For the model (5), as there are no always takers, $k = 1$. We assume $\log[p_{i3}/p_{i2}] = \theta_3 X_i$. We choose the distribution of ε from the the class of logarithmic transformations with $r = 0, 0.5$ and 1, respectively. Table 4 shows the maximum likelihood estimates, estimated standard errors and asymptotic p -values for the parameter estimates. The negative sign of $\hat{\theta}_3$ indicates that increasing age at randomization decreases the probability of being a never-taker. The three transformation models yield similar results, indicating that screening has a positive effect on survival for compliers, given the covariate (all $\hat{\beta}_2$ are negative, with p -values less than 0.01).

To select the ‘best’ model from the class of transformation models, the Akaike information criterion (AIC) could be used. In the last row of Table 4, we present the maximized log-likelihood under the three fitted transformation models. Note that the model with $r = 1$ has the highest likelihood, and consequently the lowest AIC, implying the proportional odds model fits the data best. Other criteria, for example, the Bayesian information criterion could also be used for model selection.

Figure 1 displays our estimates of $CESP(t)$ and the corresponding 95% Wald-type confidence intervals for $t = 0.5, 1, 1.5, \dots, 10$ under the proportional odds model. The dots are the point estimates and the crosses are the lower and upper limits of the confidence intervals. The triangles are the estimates of the effect of treatment assignment on survival probability $ESP(t)$ (i.e., the effect of treatment assignment without adjustment for noncompliance).

For $CESP(t)$, the 95% confidence intervals are strictly above zero after 1 year. For all t , $\hat{CESP}(t) > \hat{ESP}(t)$. Both sets of estimates are increasing in t , showing increased benefit of screening over time, with the difference between $\hat{CESP}(t)$ and $\hat{ESP}(t)$ also increasing over time. Compliers who receive annual screening have a 14.4% greater probability of surviving over 10 years than those who receive usual health care, with 95% confidence interval (0.059, 0.229). Similar results are obtained for $CESP(t; x)$ with $x = 51.56$, the sample average. For example, compliers (with $x = 51.56$) who receive annual screening have an estimated ($\hat{CESP}(10; 51.56)$) 14.5% greater probability of surviving over 10 years than those who receive usual health care, with 95% confidence interval (0.059, 0.230).

Under the Cox model, similar results are obtained: $\hat{CESP}(10)=0.147$ with 95% confidence interval (0.060, 0.234), and $\hat{CESP}(10; 51.56)=0.148$ with 95% confidence interval (0.059, 0.235).

Our results are most comparable with those obtained by Nie et al. (2011), who used the same sample selection rules. Both sets of estimates suggest that screening is beneficial, with ours suggesting a slightly higher benefit that starts earlier. As noted above, estimates obtained by other workers using different sample selection rules and methods, and in some cases different estimands, also suggest that screening is beneficial.

5. Discussion

We consider causal inference for randomized studies with right censored survival outcomes and all-or-nothing compliance, using semiparametric transformation models to estimate several causal estimands of interest. The likelihood function for the transformation models is constructed and an EM algorithm is developed to maximize the likelihood. The estimates of the model parameters are consistent and asymptotically normally distributed. Based on the assumed transformation model, we propose inferential procedures for the causal estimands. We assess the performance of the proposed procedures using a simulation study and we reanalyze the HIP data previously analyzed in a number of statistical papers.

We focus on the family of semiparametric linear transformation models, where the transformation function, aside from being monotone, is completely unspecified, while the error distribution is completely specified. This family contains as special cases the Cox proportional hazards model and the proportional odds model. The usual right censoring can be handled in a natural way and the nonparametric maximum likelihood estimator has desirable asymptotic properties. It is not difficult to see that the approach works in principle if the linear function is replaced by a nonlinear parametric function.

It would also be of interest to consider other types of semiparametric transformation models. A particularly attractive model is the one with a parametric transformation but a nonparametric additive regression function that is considered by Linton, Sperlich and Van Keilegom (2008). For survival data, it is important to handle censoring. If the censoring time is assumed to be independent of both survival time and covariates, the approach of Linton et al. (2008) may be extended through inverse probability weighting as used in Cheng et al. (1995). For causal inference with all-or-nothing compliance, further modifications may be needed to deal with mixture distributions. The semiparametric accelerated failure time (AFT) model (Kalbeisch and Prentice, 2002) is another important model in survival analysis. It would be of interest to develop approaches to these models that are computationally tractable and have solid theoretical foundation. However, in both cases above, the direct nonparametric maximum likelihood approach does not appear to work. A still different model is the semiparametric exponential tilt model for the two-sample problem that assumes a parametric form for the ratio of two underlying density functions; see Wang, Tan and Louis (2011). However, it is not clear how to extend this model to include covariates.

It would be useful to extend the results herein to studies with more than two arms. We have assumed (assumption A.7) that potential censoring times are independent of compliance type, given covariates; relaxing this assumption would allow us to make only the weaker assumption that potential event times and censoring times are independent, given observed

covariates and latent compliance type. We have also assumed all-or-nothing compliance; another useful extension would be to studies with partial compliance. Partial compliance can arise in several ways. In the JOBS-II data analyzed by Little and Yau (1998), the treatment consisted of 5 training sessions, with attendance at one or more sessions defined as taking treatment. Partial compliance also arises in longitudinal studies where the treatment consists of multiple administrations of a drug or therapy at repeated time intervals and subjects may miss one or more of the administrations. More generally, in longitudinal studies where subjects are assigned to different treatment sequences, a subject may not follow the assigned treatment in each period. A further complication occurs when treatments are allocated sequentially, depending on previous treatments, subject's covariate history and previous outcomes, necessitating consideration of time-dependent covariates. In some studies, interest focuses on whether or not treatment affects the timing of a recurrent event, for example an infection, asthma attack, or unemployment. To the best of our knowledge, although much work has been done on modeling recurrent event data (see for example Lin, Wei, Yang and Ying, 2000), non-compliance has not been taken into consideration in this context.

Acknowledgment

The authors thank the Editor, Associate Editor and two referees for their insightful comments and suggestions, which have led to many improvements. W. Yu's research was supported by the National Natural Science Foundation of China (11101091, 11271081). K. Chen's research was supported by Hong Kong RGC General Research Funds 600307, 600509, 601011 and 601612. M. Sobel's research was supported by the United States National Institutes of Health Grant RO1EB016061. Z. Ying's research was supported by the United States National Institutes of Health Grant R37GM047845.

Appendix

A.1. The EM algorithm for the proportional hazards model

When the random error ε_i in the model (5) follows the extreme value distribution, $\lambda_\varepsilon(t) = \Lambda_\varepsilon(t) = \exp(t)$, and (5) becomes the proportional hazards model. Then we have

$$\begin{aligned} l_{E,2}^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}) \\ &= \sum_{i=1}^n \sum_{k=1}^3 r_{ikd_i z_i}(\vartheta^{(0)}) \log g_{ik} \\ &= \sum_{i=1}^n \delta_i \log h(y_i) \\ &+ \sum_{i=1}^n \left\{ \delta_i \sum_{k=1}^3 l_{ik} r_{ikd_i z_i}(\vartheta^{(0)}) - H(y_i) \sum_{k=1}^3 \exp(l_{ik}) r_{ikd_i z_i}(\vartheta^{(0)}) \right\}, \end{aligned}$$

where $l_{i1} = \beta_{01} + \eta_1^\top \mathbf{x}_i$, $l_{i2} = \beta_{02} + \eta_2^\top \mathbf{x}_i$ and $l_{i3} = \beta_{03} + \eta_3^\top \mathbf{x}_i$. Here, we restrict the non-negative increasing function $H(\cdot)$ to be a right-continuous step function with possible jumps on the observed failure times. Specifically, let $y_{(1)}, \dots, y_{(n)}$ be the ordered observations and h_i the corresponding jump size at $y_{(i)}$. By replacing $H(\cdot)$ in $\mathfrak{D}$ by $(h_1, \dots, h_n)$, we obtain

$$l_{E,2}^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \vartheta, \vartheta^{(0)}) = \sum_{i=1}^n \delta_i \log h_i + \sum_{i=1}^n \left\{ \delta_i \sum_{k=1}^3 l_{ik} \cdot r_{ikd_i z_i}(\vartheta^{(0)}) - \sum_{j=1}^n I(y_j \leq y_i) h_j \cdot \sum_{k=1}^3 \exp(l_{ik}) r_{ikd_i z_i}(\vartheta^{(0)}) \right\}.$$

For fixed β_k 's and η_k 's, the h_i 's can be profiled out using

$\hat{h}_i = \delta_i / \left\{ \sum_{j=1}^n I(y_j \geq y_i) \sum_{k=1}^3 \exp(l_{jk}) r_{jkd_j z_j}(\vartheta^{(0)}) \right\}$. Plugging this expression into $l_{E,2}^O$ reduces the problem to that of maximizing

$$p_{E,2}^O(\mathcal{O}_1, \dots, \mathcal{O}_n; \beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3, \vartheta^{(0)}) \\ = \sum_{i=1}^n \delta_i \left[\sum_{k=1}^3 l_{ik} r_{ikd_i z_i}(\vartheta^{(0)}) - \log \left\{ \sum_{j=1}^n I(y_j \geq y_i) \sum_{k=1}^3 \exp(l_{jk}) r_{jkd_j z_j}(\vartheta^{(0)}) \right\} \right]$$

with respect to $(\beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3)$. $p_{E,2}^O$ can be viewed as a weighted version of the log-partial likelihood and is thus concave in the parameters $(\beta_{01}, \beta_2, \beta_{03}, \eta_1, \eta_2, \eta_3)$.

A.2. Asymptotic properties of $\hat{\vartheta}$

We prove theorem 3.1 and theorem 3.2 under suitable regularity conditions. Let $\|\cdot\|_{l^\infty[0, \tau]}$ be the supremum norm for functions on $[0, \tau]$, and $\|w\|_{BV[0, \tau]}$ the total variation of the function w on $[0, \tau]$. As the proofs are along the lines of the proof in Zeng and Lin (2006), to conform with their notation we set $G(t) = \Lambda_\varepsilon(\log t)$. We introduce the following regularity conditions.

Condition C.1. The function $H(\cdot)$ is strictly increasing and continuously differentiable on $[0, \tau]$, and ϑ_0^{par} lies in the interior of a compact set.

Condition C.2. With probability 1, $\mathbf{X}$ is bounded and not concentrated on a hyperplane of lower dimension.

Condition C.3. With probability 1, $P(Y = \tau \mid \mathbf{X} = \mathbf{x}) > 0$, where $Y = \min\{T, C\}$.

Condition C.4. $G(\cdot)$ is thrice continuously differentiable on $[0, \tau]$ with $G(0) = 0$, $G'(0) > 0$ and $\limsup_{x \rightarrow \infty} G(c_0 x)^{-1} \log(x \sup_y G'(y)) = 0$ for any positive c_0 .

Condition C.5. With probability 1, $p_k(\theta) = P(U = k \mid \mathbf{X} = \mathbf{x})$, $k = 1, 2, 3$ are uniformly bounded away from 0 and have continuous second derivatives with respect to θ . The inverse matrix of the second derivatives is uniformly bounded.

Remark: Conditions C.1–C.4 were also assumed in Zeng and Lin (2006). Condition C.4 is not satisfied if $G(\cdot)$ is a log-transformation function, but the theorems still hold with a slight modification of the proof analogous to that of Zeng and Lin (2006). Condition C.5 is a regularity condition for the parametric model of compliance type probabilities; the assumption is modified in the obvious way for the special case with no always takers.

For likelihood L^O , the log-likelihood is

$$l_n(\vartheta) = \sum_{i=1}^n [D_i Z_i \log\{\exp(\psi_{i1}) + \exp(\psi_{i2})\} + (1-D_i) Z_i \psi_{i3} + D_i (1-Z_i) \psi_{i1} + (1-D_i)(1-Z_i) \log\{\exp(\psi_{i1}) + \exp(\psi_{i2})\}]$$

where $\psi_{ik} \equiv \psi_{ik}(\vartheta)$ is the log of $p_{ik}g_{ik}$. Specifically,

$$\psi_{ik} = \Delta_i \log\{H\{Y_i\}/H(Y_i)\} + \Delta_i \log \lambda_\varepsilon\{\log H(Y_i) + \beta_{0k} + \beta_k Z_i + \eta_i^\top \mathbf{X}_i\} - \Lambda_\varepsilon\{\log H(Y_i) + \beta_{0k} + \beta_k Z_i + \eta_i^\top \mathbf{X}_i\} + \log p_{ik}(\theta)$$

and $H\{t\} = H(t) - H(t-)$.

Proof of Theorem 3.1: As in Zeng and Lin (2006), we first argue that $\hat{H}(\tau)$ must be finite.

Since condition C.4 implies $\limsup_{x \rightarrow \infty} \log \lambda_\varepsilon(x) - \Lambda_\varepsilon(x) \rightarrow -\infty$, should $H\{Y_i\}$ tend to infinity for some Y_i , $\psi_{ik} + \Delta_i \log H(Y_i)$ tends to $-\infty$. As a result, $l_n(\vartheta)$ tends to negative infinity. Hence $\hat{H}(\tau)$ must be finite. Next, we argue $\limsup_{n \rightarrow \infty} \hat{H}(\tau)$ is bounded with probability 1. Let $\bar{H}(t) = \hat{H}(t)/\hat{H}(\tau)$ and let $\bar{\vartheta}$ be the same as ϑ , except with $\hat{H}$ replaced by $\bar{H}$. Observe that $\hat{H}\{t\}/\hat{H}(t) = \bar{H}\{t\}/\bar{H}(t)$ for all $t \in [0, \tau]$. Conditions C.1–C.3 ensure there exists an $M > 0$ such that, with probability 1, $|\beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i| < M$. Then, by the boundedness of $\bar{H}$ on $[0, \tau]$ and Condition C.4,

$$\begin{aligned} \psi_{ik}(\hat{\vartheta}) - \psi_{ik}(\bar{\vartheta}) &\leq \Delta_i [\log \lambda_\varepsilon\{\hat{H}(Y_i) \\ &\quad + \beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i\} \\ &\quad - \log \lambda_\varepsilon\{\bar{H}(Y_i) \\ &\quad + \beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i\} \\ &\quad - I(Y_i \\ &\quad = \tau) \Lambda_\varepsilon\{\log \hat{H}(\tau) - M\} + \Lambda_\varepsilon(M)] \leq \Delta_i [\log \hat{H}(\tau) \\ &\quad + M + \log \left\{ \sup_{y \leq \hat{H}(\tau) + M} \lambda_\varepsilon(y) \exp(\right. \\ &\quad \quad \left. - y) \right\}] \\ &\quad - I(Y_i \\ &\quad = \tau) \Lambda_\varepsilon\{\log \hat{H}(\tau) - M\} + b \equiv \xi_i, \end{aligned}$$

where b is a constant that does not depend on k , n or i . Observe that

$$\log\{\exp(x) + \exp(y)\} - \log\{\exp(\tilde{x}) + \exp(\tilde{y})\} \leq \max(x - \tilde{x}, y - \tilde{y}).$$

It follows that

$$\begin{aligned}
& \log[\exp\{\psi_{i1}(\hat{\vartheta}) \\
& \quad + \Delta_i \log \hat{H}(Y_i)\} \\
& \quad + \exp\{\psi_{i2}(\hat{\vartheta}) \\
& \quad + \Delta_i \log \hat{H}(Y_i)\}] - \log[\exp\{\psi_{i1}(\hat{\vartheta}) \\
& \quad + \Delta_i \log \hat{H}(Y_i)\} \\
& \quad + \exp\{\psi_{i2}(\hat{\vartheta}) \\
& \quad + \Delta_i \log \hat{H}(Y_i)\}]
\end{aligned}$$

is bounded above by ξ_i . Under condition C.4, should $\hat{H}(\tau)$ diverge to infinity as $n \rightarrow \infty$, $(1/n) \sum_{i=1}^n \xi_i$ will have a subsequence diverging to $-\infty$ and so will $(1/n) \{l_n(\hat{\vartheta}) - l_n(\vartheta)\}$, contradicting the fact that $l_n(\hat{\vartheta}) - l_n(\vartheta) \rightarrow 0$. Hence, $\limsup \hat{H}(\tau) < \infty$, implying that for any subsequence, there is a further subsequence such that $\hat{\vartheta} \rightarrow \vartheta^*$. Let H^* be the nonparametric component of ϑ^* , which is a nondecreasing function on $[0, \tau]$. For simplicity of notation and without loss of generality, assume $\hat{\vartheta}$ converges to ϑ^* . We next show $\vartheta^* = \vartheta_0$, the true value of ϑ .

By differentiation, one can derive, using counting process notation $N_i(s) = \int_0^s I(Y_i \geq u) dN_i(u)$,

$$\hat{H}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{dN_i(s)}{|\phi_n(s; \hat{\vartheta})|},$$

where

$$\begin{aligned}
\phi_n(s, \vartheta) = & \frac{1}{n} \sum_{i=1}^n \left[D_i Z_i \frac{\exp\{\psi_{i1}(\vartheta)\} \zeta_{i1}(s; \vartheta) + \exp\{\psi_{i2}(\vartheta)\} \zeta_{i2}(s; \vartheta)}{\exp\{\psi_{i1}(\vartheta)\} + \exp\{\psi_{i2}(\vartheta)\}} \right. \\
& + (1 - D_i) Z_i \zeta_{i3}(s; \vartheta) + D_i (1 - Z_i) \zeta_{i1}(s; \vartheta) \\
& \left. + (1 - D_i) (1 - Z_i) \frac{\exp\{\psi_{i3}(\vartheta)\} \zeta_{i3}(s; \vartheta) + \exp\{\psi_{i2}(\vartheta)\} \zeta_{i2}(s; \vartheta)}{\exp\{\psi_{i3}(\vartheta)\} + \exp\{\psi_{i2}(\vartheta)\}} \right]
\end{aligned}$$

and

$$\zeta_{ik}(s; \vartheta) = \Delta_i \left[1 - \frac{\lambda'_\varepsilon \{\log H(Y_i) + \beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i\}}{\lambda_\varepsilon \{\log H(Y_i) + \beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i\}} \right] \frac{I(Y_i \geq s)}{H(Y_i)} + \lambda_\varepsilon \{\log H(Y_i) + \beta_{0k} + \beta_k Z_i + \eta_k^\top \mathbf{X}_i\} \frac{I(Y_i \geq s)}{H(Y_i)}.$$

The slight complication, compared with Zeng and Lin (2006), arises from the first and fourth terms in the expression for $\varphi_n(t, \vartheta)$. However, the Glivenko-Cantelli theorem still holds and $\varphi_n(t; \hat{\vartheta})$ converges to some $\varphi^*(t; \vartheta^*)$. By considering

$\tilde{H}(t) \equiv (1/n) \sum_{i=1}^n dN_i(s) / |\phi_n(s; \vartheta_0)|$, the remaining arguments carry through as in Zeng and Lin (2006). We omit the details.

Proof of Theorem 3.2: The asymptotic normality of $\hat{\vartheta}$ can be proved by citing Theorem 3.3.1 of van der Vaart and Wellner (1996). View $\hat{\vartheta}$ and ϑ as elements of the Banach space $\mathbb{R}^m \times l^\infty(\mathcal{D})$. For any ϑ , $\Psi(\vartheta)$ in the theorem is the expectation of the score, which is also an element in the same Banach space, with empirical analogue $\Psi_n(\vartheta)$. Then, $\Psi(\vartheta_0) = 0 = \Psi_n(\hat{\vartheta})$. To verify Ψ , the Frechet derivative of Ψ at the ϑ_0 , as a linear mapping from the linear space spanned by the domain of ϑ to the Banach space, is continuously invertible, one can consider any one-dimensional parametric submodel $\vartheta_0 + a(\tilde{\vartheta} - \vartheta)$, with a being the one-dimensional parameter. The Fisher information for this parametric model at $a = 0$ is bounded away from 0 uniformly over all $\tilde{\vartheta}$ with unit norm for $\tilde{\vartheta} - \vartheta$. This ensures the invertibility of Ψ and the theorem implies the asymptotic normality of $\hat{\vartheta}$.

References

- Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J. Am. Statist. Ass.* 1996; 91:444–455.
- Baker SG. Analysis of survival data from randomized trial with all-or-none compliance: Estimating the cost-effectiveness of a cancer screening program. *J. Am. Statist. Ass.* 1998; 93:929–934.
- Baker SG. Estimation and inference for the causal effect of receiving treatment on a multinomial outcome: An alternative approach. *Biometrics.* 2011; 67:319–325. [PubMed: 20560933]
- Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statist. Med.* 1994; 13:2269–2278.
- Bloom HS. Accounting for no-shows in experimental evaluation designs. *Evaluation Review.* 1984; 8:225–246.
- Chen K, Jin Z, Ying Z. Semiparametric analysis of transformation models with censored data. *Biometrika.* 2002; 89:659–668.
- Cheng J. Estimation and inference for the causal effect of receiving treatment on a multinomial outcome. *Biometrics.* 2009; 65:96–103. [PubMed: 18373714]
- Cheng J, Qin J, Zhang B. Semiparametric estimation and inference for distributional and general treatment effects. *J. R. Statist. Soc. B.* 2009a; 71:881–904.
- Cheng J, Small D, Tan Z, Ten Have T. Efficient nonparametric estimation of causal effects in randomized trials with noncompliance. *Biometrika.* 2009b; 96:19–36.
- Cheng SC, Wei LJ, Ying Z. Analysis of transformation models with censored data. *Biometrika.* 1995; 82:835–845.
- Cuzick J, Edward R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Statist. Med.* 1997; 16:1017–1029.
- Cuzick J, Sasieni P, Myles J, Tyrer J. Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *J. R. Statist. Soc. B.* 2007; 69:565–588.
- Frangakis CR, Rubin DB. Addressing complications of intent to treat analysis in the presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika.* 1999; 86:365–379.
- Gong, Z. PhD Dissertation. Christchurch: University of Canterbury; 2008. Parametric potential-outcome survival models for causal inference.
- Hirano K, Imbens GW, Rubin DB, Zhou XH. Assessing the effect of an influenza vaccine in an encouragement design. *Biostatistics.* 2000; 1:69–88. [PubMed: 12933526]
- Imbens GW, Angrist JD. Identification and estimation of local average treatment effects. *Econometrica.* 1994; 62:467–476.
- Imbens GW, Rubin DB. Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies.* 1997; 64:555–574.
- Jin H, Rubin DB. Principal stratification for causal inference with extended partial compliance. *J. Am. Statist. Ass.* 2008; 103:101–111.

- Joffe MM. Administrative and artificial censoring in censored regression models. *Statist. Med.* 2001; 20:2287–2304.
- Kalbeisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. New York: Wiley; 2002.
- Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. *J. R. Statist. Soc. B.* 2000; 62:711–730.
- Linton O, Sperlich S, Van Keilegom I. Estimation of a semiparametric transformation model. *Ann. Statist.* 2008; 36:686–718.
- Little RJ, Yau LHY. Statistical techniques for analyzing data from prevention studies: Treatment of no-shows using Rubin's causal model. *Psychological Methods.* 1998; 2:147–159.
- Loeys T, Goetghebeur E. A causal proportional hazards estimator for the effect of treatment actually received in a randomized trial with all-or-nothing compliance. *Biometrics.* 2003; 59:100–105. [PubMed: 12762446]
- Mark SD, Robins JM. Estimating the causal effect of smoking cessation in the presence of confounding factors using a rank preserving structural failure time model. *Statist. Med.* 1993; 12:1605–1628.
- Nie H, Cheng J, Small D. Inference for the effect of treatment on survival probability in randomized trials with noncompliance and administrative censoring. *Biometrics.* 2011; 67:1397–1405. [PubMed: 21385167]
- Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Comm. Statist.-Theor. Meth.* 1991; 20:2609–2631.
- Rubin D. Comment on 'Randomization analysis of experimental data: the Fisher randomization test' by D. Basu. *J. Am. Statist. Ass.* 1980; 75:591–593.
- Shapiro, S.; Venet, W.; Strax, P.; Venet, L. *Periodic Screening for Breast Cancer: The Health Insurance Plan Project and Its Sequelae, 1963–1986*. Baltimore: The Johns Hopkins University Press; 1988.
- Small D, Ten Have T, Joffe MM, Cheng J. Random effects logistic models for analysing efficacy of a longitudinal randomized treatment with non-adherence. *Statist. Med.* 2006; 25:1981–2007.
- Sobel ME, Muthen B. Compliance mixture modelling with a zero effect complier class and missing data. *Biometrics.* 2012; 68:1037–1045. [PubMed: 22985224]
- Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Statist. Med.* 1991; 10:45–52.
- van der Vaart, AW.; Wellner, JA. *Weak Convergence and Empirical Processes: with Application to Statistics*. New York: Springer-Verlag; 1996.
- Wang C, Tan Z, Louis TA. Exponential tilt models for two-group comparison with censored data. *J. Statist. Plann. and Inference.* 2011; 141:1102–1117.
- Zeng D, Lin DY. Efficient estimation of semiparametric transformation models for counting processes. *Biometrika.* 2006; 93:627–640.
- Zeng D, Lin DY. Maximum likelihood estimation in semi-parametric regression models with censored data (with discussion). *J. R. Statist. Soc. B.* 2007; 69:507–564.

Figure 1.

Results from the HIP study under the proportional odds model. We estimate $CESP(t)$ and $ESP(t)$ for $t = 0.5, 1, 1.5, \dots, 10$. The dots are the point estimates of the complier effects. The crosses are the lower and upper limits of the corresponding 95% Wald-type confidence intervals. The triangles are the estimates of the effects of treatment assignment without adjustment for noncompliance.

Table 1

Summarized simulation results for the parameters of the three models.

Model	Parameter	True Value	$n = 100$					$n = 200$				
			AVE	SE	SEE	CP	AVE	SE	SEE	CP	SE	CP
$r = 0$	θ_1	1	1.0702	0.3921	0.3620	0.960	1.0277	0.2609	0.2459	0.941		
	θ_3	-1.2	-1.2963	0.4830	0.4300	0.968	-1.2472	0.3041	0.2896	0.961		
	β_{01}	-0.5	-0.5995	0.8532	0.8023	0.925	-0.5308	0.5314	0.5057	0.941		
	β_2	-2	-2.2011	0.6432	0.6073	0.962	-2.1138	0.3957	0.3774	0.959		
	β_{03}	1	1.0834	0.8717	0.8961	0.940	1.0029	0.6174	0.5598	0.933		
	η_1	0.3	0.4097	0.7301	0.6449	0.910	0.3229	0.4252	0.3980	0.930		
	η_2	0	-0.0356	0.3553	0.3524	0.957	-0.0254	0.2415	0.2295	0.947		
	η_3	-0.3	-0.2941	0.7672	0.7468	0.933	-0.2658	0.4891	0.4549	0.940		
	θ_1	1	1.0487	0.3940	0.3641	0.948	1.0338	0.2604	0.2496	0.953		
	θ_3	-1.2	-1.2880	0.4907	0.4342	0.960	-1.2284	0.2919	0.2879	0.955		
	β_{01}	-0.5	-0.5518	1.1393	1.0945	0.945	-0.4794	0.7369	0.7118	0.931		
	β_2	-2	-2.2347	0.7342	0.6763	0.956	-2.0913	0.4517	0.4249	0.949		
	β_{03}	1	1.1341	1.3597	1.2662	0.910	1.0532	0.8514	0.8089	0.930		
$r = 0.5$	η_1	0.3	0.4194	0.9238	0.8125	0.929	0.3253	0.5612	0.5250	0.944		
	η_2	0	0.0026	0.5381	0.4984	0.930	0.0174	0.3322	0.3224	0.949		
	η_3	-0.3	-0.2817	1.2615	0.9959	0.905	-0.2507	0.6630	0.6055	0.930		
	θ_1	1	1.0481	0.3937	0.3656	0.943	1.0353	0.2613	0.2511	0.954		
	θ_3	-1.2	-1.2877	0.4903	0.4322	0.958	-1.2258	0.2908	0.2878	0.958		
	β_{01}	-0.5	-0.5443	1.4475	1.3526	0.927	-0.4981	0.9305	0.8919	0.930		
	β_2	-2	-2.2303	0.8553	0.7613	0.947	-2.0799	0.5093	0.4911	0.945		
	β_{03}	1	1.1542	1.7699	1.5425	0.915	1.0391	1.0852	1.0094	0.930		
	η_1	0.3	0.4023	1.0946	0.9637	0.922	0.3293	0.6869	0.6314	0.933		
	η_2	0	0.0080	0.6678	0.6062	0.934	0.0154	0.4097	0.3993	0.948		
	η_3	-0.3	-0.3162	1.5681	1.2058	0.908	-0.2500	0.8238	0.7349	0.925		

AVE: average of the estimates; SE: standard error of the estimates; SEE: average of the estimated standard errors; CP: empirical coverage probabilities of Wald-type confidence intervals with 95% confidence level.

Table 2

Summarized simulation results for the time t effect on the complier survival probability (CESP) for the three models.

Estimand	Model	t	True Value*	$n = 100$				$n = 200$			
				AVE	SE	SEE	CP	AVE	SE	SEE	CP
CESP($t;1$)	$r = 0$	1	0.5055	0.5081	0.1043	0.1004	0.922	0.5109	0.0694	0.0674	0.937
		2	0.6275	0.6384	0.1115	0.1100	0.920	0.6378	0.0766	0.0751	0.940
		3	0.6165	0.6208	0.1205	0.1229	0.930	0.6248	0.0855	0.0849	0.936
		4	0.5637	0.5599	0.1281	0.1277	0.921	0.5680	0.0907	0.0891	0.944
	$r = 0.5$	1	0.4328	0.4342	0.1186	0.1137	0.910	0.4293	0.0804	0.0766	0.931
		2	0.5258	0.5390	0.1255	0.1212	0.924	0.5294	0.0863	0.0830	0.941
		3	0.5310	0.5489	0.1322	0.1268	0.920	0.5379	0.0907	0.0877	0.932
		4	0.5082	0.5277	0.1354	0.1303	0.915	0.5165	0.0929	0.0901	0.936
	$r = 1$	1	0.3808	0.3796	0.1258	0.1180	0.920	0.3777	0.0862	0.0818	0.930
		2	0.4537	0.4621	0.1353	0.1293	0.924	0.4552	0.0931	0.0894	0.935
		3	0.4612	0.4748	0.1406	0.1344	0.922	0.4650	0.0957	0.0929	0.935
		4	0.4488	0.4650	0.1429	0.1369	0.910	0.4537	0.0962	0.0942	0.940
CESP(t)	$r = 0$	1	0.5055	0.5070	0.0984	0.0986	0.927	0.5029	0.0680	0.0662	0.939
		2	0.6275	0.6309	0.1109	0.1090	0.914	0.6343	0.0763	0.0747	0.938
		3	0.6165	0.6139	0.1217	0.1209	0.914	0.6219	0.0849	0.0842	0.936
		4	0.5637	0.5571	0.1237	0.1246	0.920	0.5662	0.0897	0.0880	0.940
	$r = 0.5$	1	0.4328	0.4320	0.1139	0.1100	0.917	0.4285	0.0788	0.0751	0.934
		2	0.5258	0.5321	0.1229	0.1192	0.930	0.5263	0.0859	0.0825	0.940
		3	0.5310	0.5409	0.1294	0.1251	0.924	0.5342	0.0905	0.0871	0.934
		4	0.5082	0.5201	0.1317	0.1280	0.920	0.5129	0.0923	0.0891	0.936
	$r = 1$	1	0.3808	0.3771	0.1215	0.1152	0.923	0.3765	0.0845	0.0802	0.933
		2	0.4537	0.4556	0.1325	0.1280	0.930	0.4519	0.0924	0.0887	0.937
		3	0.4612	0.4672	0.1380	0.1323	0.923	0.4613	0.0952	0.0923	0.935
		4	0.4488	0.4575	0.1397	0.1341	0.920	0.4502	0.0955	0.0933	0.933

* CESP($t;1$)=CESP(t) for all t as $\eta_2 = 0$.

AVE: average of the estimates; SE: standard error of the estimates; SEE: average of the estimated standard errors obtained by the delta method; CP: empirical coverage probabilities of Wald-type confidence intervals with 95% confidence level.

Table 3
Summarized simulation results for the complier median causal effects for the proportional hazards model.

Estimand	True value*	n = 100			n = 200		
		AVE	SE	SEE	AVE	SE	CP
CQCE(0.5;1)	4.4286	4.5780	1.8595	1.8330	4.2090	1.1678	0.949
CQCE(0.5)	4.4286	4.5601	1.8326	1.8253	4.5814	1.3838	0.951

* CQCE(0.5;1)=CQCE(0.5) as $\eta_2 = 0$.

AVE: simulated average of the estimates; SE: simulated standard error of the estimates; SEE: average of the estimated standard errors obtained by the bootstrap method; CP: empirical coverage probabilities of Wald-type confidence intervals with 95% confidence level.

Table 4

Results from the HIP study under linear transformation models: maximum likelihood estimates of the model parameters.

Parameters	$r = 0$			$r = 0.5$			$r = 1$		
	EST	SE	p -Value	EST	SE	p -Value	EST	SE	p -Value
θ_3	-0.0179	0.0021	<0.0001	-0.0179	0.0021	<0.0001	-0.0179	0.0021	<0.0001
β_2	-0.5350	0.1593	0.0008	-0.5911	0.1759	0.0007	-0.6456	0.1925	0.0008
β_{03}	-1.7228	1.5394	0.2617	-1.9757	1.6756	0.2384	-2.5924	1.8084	0.1486
η_2	-0.0101	0.0117	0.3839	-0.0123	0.0131	0.3464	-0.0159	0.0143	0.2605
η_3	0.0176	0.0235	0.4523	0.0202	0.0255	0.4286	0.0286	0.0276	0.2984
Log-likelihood		-2496.49			-2493.13			-2492.79	

EST: maximum likelihood estimate; SE: estimated standard error; p -value: asymptotic p -value of the significance test.